

Rohit Kumar

Meerut Uttar Pradesh • rohit2852001@gmail.com • +91 9759279921 • [LinkedIn](#) • <https://github.com/git-RohitHub>

• [Website/Portfolio](#)

EDUCATION

DEWAN V.S INSTITUTE OF ENGINEERING AND TECHNOLOGY , Meerut Uttar Pradesh
B.Tech in Computer Science

Jan 2020 - Jan 2024
GPA: 7.3

Relevant coursework: Full Stack Data Science (Physics Wallah) , Data Analytics (ICT Academy) , Big Data Analytics (CDAC NASSCOM)

PROFESSIONAL EXPERIENCE

Frisson Devhub - Noida Uttar Pradesh
AI/ML Engineer

Nov 2023 - Present

- Developed scalable data pipelines, processing 10,000+ rows daily to ensure data accuracy and consistency.
- Designed machine learning models with TensorFlow and PyTorch, achieving 90% accuracy in classification tasks.
- Built model backends with FastAPI and Flask, improving latency by 25% and enhancing application performance. Automated model deployment workflows by utilizing CI/CD pipelines and MLOps techniques, reducing deployment time by 30% and ensuring robust production environments.
- Implemented Retrieval-Augmented Generation (RAG) in Generative AI solutions, such as chatbots, by integrating tools like LangChain, LangGraph, Vector DB, and LLMs, resulting in a 20% improvement in response relevance and accuracy.

PROJECTS

SIFRA: Empowering Secure and Customizable Chatbots for Businesses

Nov 2024 - Jan 2025

- Description: Developed a platform enabling businesses to create secure, private chatbots trained on proprietary data, integrated seamlessly into websites and communities.
- Tools & Technologies: Python, BeautifulSoup, LangChain, LLMs, Vector DB, Fast API, AWS
- Designed an ETL pipeline to preprocess 10,000+ records for chatbot training.
- Built a high-accuracy vectorstore using LangChain for embedding storage and retrieval.
- Researched and optimized base models, improving chatbot performance by 15%.
- Engineered prompts for LLMs, enhancing response relevance by 20%.
- Evaluated retrievers (e.g., Naïve, Parent) for optimal context retrieval.
- Develop GitHub integration for automated documentation updates.

AI Text Detection: Identifying Human, AI, and Hybrid Content

Sep 2024 - Nov 2024

- Description: Developed a platform to classify written content as human-written, AI-generated, or hybrid. Applications include plagiarism detection, content moderation, and AI usage monitoring in professional writing.
- Tools & Technologies: Python, Hugging Face, LLMs , Data Cleaning, Data Preparation, Pytorch, AWS Sagemaker, Fast API
- Data Collection: Curated diverse datasets, including human-written texts from communities, AI-generated content from GPT, Claude, and Copilot, and hybrid samples.
- Data Annotation: Labeled data with precise categories for supervised learning, ensuring high-quality training inputs.
- Data Preparation: Designed structured input-output formats to enable seamless integration with LLMs and accurate predictions.
- Fine-Tuning: Trained LLaMA 3.1 8B, achieving 94% training accuracy and 89% testing accuracy in classification.
- Model Deployment: Published the fine-tuned model on Hugging Face and deployed it on AWS SageMaker with Lambda, enabling scalable and efficient inference

Business Email Generator: Personalized Job Application Emails

Dec 2024 - Dec 2024

- Description: Created a system to generate personalized job application emails by analyzing job postings and matching them with user profiles.
- Tools: LangChain, Python, Groq, Chroma (Vector DB), LLaMA 3.1 70B, Streamlit, Flask
- Scraped job postings in real-time and stored details in a vectorstore.
- Integrated user skills and portfolio for precise email generation.
- Automated email creation using job URL as input.

SKILLS

Functional : Agile, JIRA, CI/CD
Technical : Python, Machine Learning, Deep Learning, NLP, Pytorch, Keras, EDA, Data Analysis, AWS Sagemaker, Langchain, Phidata, RAG, LLM, Flask, Fast API, SQL, Power BI, Docker, Git, AWS, Vector DB
Data Analytics : Numpy, Pandas, Matplotlib, Seaborn, Plotly